
MEMORANDUM

P.O. Box 2393
Princeton, NJ 08543-2393
Telephone (609) 799-3535
Fax (609) 799-0005
www.mathematica-mpr.com

TO: Paul O'Leary

FROM: Abigail Philip, Cassie Pickens Jewell, Michelle Lee, and Janine Bologna

DATE: 5/18/2018

SUBJECT: DAF16 Documentation Deliverables Summary

This memo summarizes the updates made to DAF16 documentation, which includes Volumes 1 – 11 and the Code Library. Each year, the DAF documentation team updates over 3,500 pages of text, graphs, and tables to reflect changes to the DAF. In addition to the usual year-over-year changes, the team also addresses database enhancements, which are prioritized each year through the Assess Data Needs process. This memo focuses on: 1) Common year-over-year changes; 2) DAF16 enhancements, highlighting any major changes to the documentation; and 3) Volume 1 – 3, responses to comments.

A. YEAR-OVER-YEAR CHANGES: Year-over-year changes are the changes that the documentation team expects to make every year. A few examples of year-over-year changes are the following: updating references to the DAF year (i.e. DAF15 becomes DAF16), updating record counts, and reviewing footnotes. Year-over-year changes were made in all 11 volumes and in the code library, as appropriate.

B. VOLUME-SPECIFIC ENHANCEMENT UPDATES: Volume-specific enhancements encompass revisions that are a direct result of major changes to the database each year. Generally speaking, each major change to the DAF database results in some associated text or explanation in one or more of the volumes. Below is a list of the major changes by volume.

Volume 1: Getting Started with the DAF16

- Updated Chapter II, Section E.2 and Appendix A, which details the selection criteria for DAF, to include the following beneficiaries:
 - All beneficiaries from 18 through FRA.
 - Children who received SSI benefits under the age of 10 in at least one month since 2005.

MEMO TO: Paul O'Leary

FROM: Abigail Philip, Cassie Pickens Jewell, Michelle Lee, and Janine Bologna

DATE: 5/18/2018

PAGE: 2

- Updated Chapter II, Section E.3 to remove the following text: “*approximately 2 percent of records for SSDI auxiliary beneficiaries are excluded.*” This statement was replaced with text that describes that 1) some of these cases might get missed, 2) that this is due to an problem with dependents (mostly DACs) not having their own SSN, and 3) that we expect this problem to be small and shrinking because of enumeration at birth and other efforts to have BOANs for all beneficiaries.
- Updated Figure III.1 (Diagram of the DAF) to remove Non-Enrolled Annuals files and include the two DAF-linkable data files, RSA-911 files and the MEF, which are also available on the SSA mainframe.
- Revised Chapter III, Section B to omit text on the Non-Enrolled Annual Files.

Volume 2: Working with the DAF16

- Replaced Chapter III (Understanding the Meaning of Blank or Missing Variable Values in the DAF) and Chapter IV (Variables in the DAF DMG Component File from the 831 & 832/833 Files) with a new chapter titled “The Importance of the Administrative Source for Understanding How to Use DAF Data.” This section includes information on CER/DBAD, which came from the PUF. To improve the flow of this section, we also moved some MBR information into this section. Additionally, we moved the “Other missing values” section, which was originally found in Chapter III, to Chapter II and renamed it “Missing values in the DAF.” We also updated this section of the text to include some clarifying language on the differences between missing, blank and zero-filled variables.
- Updated Figure V.1 (Layout of the DAF16 10% data mart components) to remove Non-Enrolled Annuals files and include the two DAF-linkable data files, RSA-911 files and the MEF, which are also available on the SSA mainframe.
- Updated Table V.1 (Number and percent of records in the DAF16 10% data mart) to note the removal of beneficiaries from the Non-Enrolled Annual files.

Volume 3: Tips for Conducting Analysis with the DAF16

- Updated Chapter II, Section A with text on the addition of two new variables, LAF_MR and LAF_MR_DT, to the DMG.
- Updated Chapter IV, Section A with the following:
 - Added text describing SSI income and related exclusions and it being sourced from the SSR instead of the DCF.

MEMO TO: Paul O'Leary

FROM: Abigail Philip, Cassie Pickens Jewell, Michelle Lee, and Janine Bologna

DATE: 5/18/2018

PAGE: 3

- Created Table IV.1. SSI earnings measures from various SSA administrative data files, as named in the DAF, which identifies the SSI earnings variables from the CER100% file, the SSR, and the DCF, and provides the DCF variables for users more familiar with using them, either directly from the DCF or from earlier versions of the DAF.
 - Added text describing the new summary measure SSIERNLVlyymm that indicates the earnings level of SSI beneficiaries, derived from the SSR.
- Updated Chapter IV, Section B by accepting client's track changes in this section and adding text that (1) describes how earnings for SSI are applied based on when paid, while for SSDI they are applied based on when earned and (2) the changes that resulted from the Bipartisan Budget Act of 2015 and (3) earnings recorded for SSDI beneficiaries may not always be reliable or consistent with IRS annual earnings data. We noted that the new policy that resulted from BBA Section 825 does not really change this distinction, but gives SSA the flexibility to use when paid for SSDI if information on when earned is not "readily available." This means that after September 23, 2016, the application of earnings for SSDI will be a mix of when paid and when earned.
- Updated Chapter VII to make the language more current (it previously was only applicable to DAF15, but we made a few tweaks to make it less year specific) and moved some text around to improve readability.
- Updated Chapter VIII to omit the comparison of DAF14 to DAF15 SCWF and to note that calculations used DAF16 values.
- Added references to Code Library throughout the volume.

Volume 4: Lists of DAF16 Variables

- Updated language for Chapter II, Section L (Variables related to earnings, income, and work) to incorporate update information about the SSI earning values sourced from the SSR, as well as information about SSR about self-employment income (T16SEAMT) and blind work expense (T16BEXPAMT).
- Updated all tables in Chapters II and III to include detailed information about DAF Component (e.g., Payment Horizontal, Payment Vertical, Ticket Annual, Ticket Base).

MEMO TO: Paul O'Leary
FROM: Abigail Philip, Cassie Pickens Jewell, Michelle Lee, and Janine Bologna
DATE: 5/18/2018
PAGE: 4

Volume 5: DAF Variable Detail Pages¹

- Added detailed variable values and SAS code information to all 289 RSA variables.
- Made variable label, variable description, and variable name changes, as appropriate, to reflect the preferred “SVR Agency” language (rather than VR) to more than 70 variables.
- Provided more detailed information about variables components (e.g., Payment Horizontal, Payment Vertical, Ticket Annual, Ticket Base) for 202 variables.
- Reviewed, revised, and updated variable information related to “Data For,” “Data Sources,” “Time Period,” and “Program Name” fields for all 788 variables.
- Revised “Notes” language for seven IEA_variables.
- Updated/added administrative data language or SAS Code field revisions for 499 non-RSA variables (e.g., DIRPAYyymm=Extracted from PHUS (DISPAY)).

Volume 6: Validating the DAF16 Against Other Sources

- After comparing data from DAF16 with data from the SSA published tables, the documentation team updated Volume 6 appendices, which include a series of tables and associated graphs for DAF16, DAF16 10%, and the ADM.

Appendix	Title	Number of updated tables	Number of updated graphs
Appendix A	DAF16	20	20
Appendix B	DAF16 10%	25	25
Appendix C	ADM	2	2

Volume 7: DAF16 Development History and Construction Methods

- Adjusted references to the companion Non-enrolled Annual file to reflect that the production of this document was discontinued for DAF16.
- Updated Section II, “Overview of DAF construction tasks,” to reflect changes to DAF16 programming. For example, the documentation team updated Task 7, “Process SSI-LF

¹ This list comprises major changes made to the variable detail pages. Additional changes, such as revisions to individual variables based on comments from programmers and/or SSA, are not included in this list.

MEMO TO: Paul O'Leary

FROM: Abigail Philip, Cassie Pickens Jewell, Michelle Lee, and Janine Bologna

DATE: 5/18/2018

PAGE: 5

data” to describe how we use two custom extracts of select SSR variables to create one record per SSN with monthly variables.

Volume 8: DAF16 Construction Workflow Charts and Task Tables

- No major updates were made to Volume 8.

Volume 9: DAF16 Source File Descriptions

- No major updates were made to Volume 9.

Volume 10: DAF Administrative Source File Documentation

- No major updates were made to Volume 10.

Volume 11: DAF16 Construction Code and Data Mart Details

- Moved the “Documentation for Processing of RSA Data with DAF Data memo” from Volume 11, Appendix B to Volume 2, Appendix D.
- Moved the “MEF Documentation memo” from Volume 11, Appendix D to Volume 2, Appendix E.

All DAF16 Volumes:

- To make searching for mentions of the DAF easier, we included a suggested citation in the documentation so that disability researchers may correctly cite the DAF. The suggested citations are located on the back of the cover page of each volume.

C. RESPONSE TO COMMENTS: VOLUMES 1-3, & 7: The section below is a bulleted list of client feedback on draft DAF16 documentation (Volumes 1-3). In addition, we addressed two comments from the DAF15 documentation review (Volumes 2 and 7), which is noted with footnotes.

Volume 1: Getting Started with the DAF16

- ***“The Non-Enrolled Annuals files, which tracked data for beneficiaries in years that they were not receiving benefits, are no longer being included in the DAF. These files were of limited research use and have been omitted to make the file size of the DAF more manageable.”*** Please provide a bit more explanation in terms of explaining that these observations include only missing data because there was no activity in the year because neither the SSI nor the SSDI programs were active in the year for these individuals.

MEMO TO: Paul O'Leary

FROM: Abigail Philip, Cassie Pickens Jewell, Michelle Lee, and Janine Bologna

DATE: 5/18/2018

PAGE: 6

- The text has been updated to say the following:

“The Non-Enrolled Annuals files, which tracked data for beneficiaries in years that they were not receiving benefits, are no longer being included in the DAF. These files contained records for beneficiaries in years they were not receiving benefits. As such, they were of limited research use because they contained only missing data. Because they were not useful for research, they have been omitted to make the file size of the DAF more manageable.” (p. 8)

- ***“We have added a new summary measure SSIERNLVlyymm that indicates the earnings category level of SSI beneficiaries, derived from the Supplemental Security Record (SSR). This measure is similar to the variable added in DAF15 for SSDI beneficiaries. We describe the SSI and SSDI measures in Volume 3.”*** Note also that we added the split variables for Self-employment and Net loss from self-employment as well as Blind work expense (Blind IRWE) and Disabled workers impairment-related work expenses (IRWE).

- We believe this suggestion fits better with the bullet point about replacing information related to SSI income and related exclusions, to be sourced from the SSR instead of the Disability Control File (DCF). The text has been updated to say the following:

“We have replaced information related to SSI income and related exclusions, to be sourced from the SSR instead of the DCF. While these variables largely aligned in the DCF and SSR, there were certain cases where values in the DCF were missing. As such, we are now sourcing these variables from the more reliable measures in the SSR. We describe this change in more detail in Volume 3. By changing the source of these variables, we were also able to obtain a measure of blind work expenses separated from disabled workers work expenses, as well as self-employment income separated from the net loss from self-employment income.” (p. 8)

- **Table II.1 Years that beneficiary age groups were first selected for inclusion in the DAF. This whole table is confusing and needs to be reworked. We first selected those age 10-18 starting in 2005. And it is confusing to say SSDI and SSI beneficiaries ages 65 to FRA were included in the DAF in 2007, especially given the text above, “For example, when we first selected SSI and SSDI beneficiaries between the ages of 65 and FRA, we selected cases from 2007 onward. In DAF16, were re-selected beneficiaries in those ages from earlier years and now include them in the DAF.” Please clarify.**

- The text above the table has been updated to say the following:

MEMO TO: Paul O'Leary
FROM: Abigail Philip, Cassie Pickens Jewell, Michelle Lee, and Janine Bologna
DATE: 5/18/2018
PAGE: 7

“As mentioned above and described in detail in Volume 7, we apply specific selection criteria when selecting new beneficiaries in the current DAF year. Beneficiaries selected for inclusion in earlier DAF iterations were subject to the selection criteria in place at that time. In DAF16, we expanded the selection criteria to include beneficiaries that we did not capture with our prior selection criteria. As a result, the DAF16 currently contains the universe of beneficiaries who were received SSDI in any year from 1996 through 2016, and were ages 18 to FRA in that same year; beneficiaries who received SSI in any year from 1996 through 2004 and were ages 18 to FRA in that year; and beneficiaries who received SSI in any year from 2005 through 2016 and were under FRA in that year. In other words, all SSDI beneficiaries are now included in the DAF, and all disabled SSI beneficiaries from 2005 onward are in the DAF. Prior to 2005, SSI beneficiaries included in DAF16 are limited to adults; we intend to add children for 1996 through 2004 in a future version of the DAF.” (p. 11)

- The table and accompanying footnotes have also been updated accordingly:

- **Table II.1 Years that beneficiary age groups are selected for inclusion in the DAF**

Beneficiary group	Years selected for inclusion in the DAF... ¹
SSDI beneficiaries ages 18 to FRA ²	1996 onward
SSI beneficiaries ages 18 to FRA	1996 onward
SSI beneficiaries ages 0 to 18	2005 onward

¹ Note, the year shown for inclusion reflects the years in which beneficiaries could have been included in the DAF based on the age selection criteria. This does not indicate the DAF version that we started to include those beneficiaries.

² See Volume 3 for more information about how FRA has changed over time.

Volume 2: Working with the DAF16

- **For the DAF15 we have added links to these memos to our documentation page. If we incorporate these memos into the documentation, the obvious candidate would be to include them in Volume 2 since that volume already includes sections on linkable files and extracts.²**
 - The following memos have been added as appendices in Volume 2:
 - Appendix A: DAF16 Ticket to Work Linkable File Extract
 - Appendix B: DAF16 National Beneficiary Survey (NBS) Linkable File Extract

² This comment is from the client's review of DAF15 documentation and was addressed in DAF16.

MEMO TO: Paul O'Leary

FROM: Abigail Philip, Cassie Pickens Jewell, Michelle Lee, and Janine Bologna

DATE: 5/18/2018

PAGE: 8

- Appendix C: DAF16 Survey and Demonstration Projects Linkable File Extract
- Appendix D: DAF16 Rehabilitation Services Administration (RSA) Linkable File
 - Note: Moved from Volume 11.
- Appendix E: DAF16 Master Earnings File (MEF) Documentation and JCL/SAS Code
 - Note: Moved from Volume 11. Also, renamed from MEF Documentation to MEF Documentation and JCL/SAS Code.

Volume 3: Tips for Conducting Analysis with the DAF16

- ***“All of these documents are available to SSA staff on the SSA intranet at <https://www.ssa.gov/disabilityresearch/daf.html>.” This is now the public internet site rather than the intranet. Please check and correct this in the other volumes as well.***
 - We have removed “to SSA staff on the SSA intranet” in the sentence and have made this change to all the volumes. (p. xi)
- **Table IV.1. SSI earnings measures from various SSA administrative data files, as named in the DAF. T16EXPAMTyymm should be merged to cover BEXP and EXP. The DCF combined these variables as simply an “expense amount” and so lost details in the SSR source data. By using the SSR, we can now recover those details. We should add a similar footnote as “2,” something like: The T16EXPAMTyymm variable contained in the DCF was a composite measure of Blind work expenses (T16BEXPAMTyymm) and non-blind impairment-related work expenses (IRWE—T16EXPAMTyymm); beneficiaries in the SSR may have one, though not both, of those values populated in a given month.**
 - We have added the suggested footnote to the T16EXPAMTyymm variable. (p. 17)
- **Table VIII.2. A comparison of STWSSI and STWSSI_FCI in January 2013. I do not think updating this table is necessary for DAF16. Just add a sentence to the “Note” saying we do not expect these statistics would have measurably changed for the DAF16, so we have not updated them.**
 - Because we updated Table VIII.6 using DAF16 values, to maintain consistency, we updated this table using DAF16 values as well. The table has been updated to say the following (p. 54):

MEMO TO: Paul O'Leary

FROM: Abigail Philip, Cassie Pickens Jewell, Michelle Lee, and Janine Bologna

DATE: 5/18/2018

PAGE: 9

Table VIII.2. A comparison of STWSSI and STWSSI_FCI in January 2013

1.	2. Total	3. STWSSI_FCI1301 (row percent)						
STWSSI1301		0	1	2	3	4	8	9
0	5,279,383	100.0	0.0	0.0	0.0	0.0	0.0	0.0
1	72,941	0.0	97.6	0.0	0.1	1.8	0.5	0.0
2	495	0.0	0.2	79.0	1.6	0.2	5.5	13.7
3	50,786	0.0	0.0	0.0	81.4	0.0	0.0	18.6
4	32,119	0.0	0.3	0.0	0.01	99.5	0.05	0.0
8	1,186,352	0.0	0.0	0.0	0.15	0.0	99.85	0.0
9	3,606,593	0.0	0.0	0.0	0.15	0.0	0.0	99.85

Note: The STWSSI_FCI1201 values are the percentage of the total shown in the first column. These statistics were developed using the DAF16. We selected January 2013 to allow any lags in recording of variables used to generate STWSSI to be minimal.

- ***“Though the data is the same in each case, not all of the columns are used in each calculation, and as such, we gray out columns that were not necessary for the relevant BFWSSI calculation.”*** I still find the greyed columns useful so I would recommend retaining them.
 - We were initially unsure if we should omit the grey columns, especially since it details the process of the change from DAF14 to DAF15 and may not be applicable to what we need now. But per your recommendation, we will leave in the grey columns. (p. 58)
- **Table VIII.6. A comparison of the new and old BFWSSI measures, January 2013. I would like to update this table to use the 2016 data. I say this only because we use 2016 in the walkthrough above and, since this shows what the implication of that walkthrough is, I think it best to use 2016 data here.**
 - The table has been updated with DAF16 values (p. 62):

Table VIII.6. A comparison of the new and old BFWSSI measures, January 2013

BFWSSI_DRAFT1301	SSI Beneficiaries with BFWSSI_DRAFT in bin	BFWSSI_DRAFT_BOTH1301 compared to BFWSSI_DRAFT1301 (row percent)		
		Less	Same	More
\$0 ¹	18,770,883	NA	99.9	0.0
Under \$10	10,724	8.5	90.7	0.8
\$10 to \$24	19,417	7.9	91.4	0.8
\$25 to \$49	32,351	9.8	89.7	0.4

MEMO TO: Paul O'Leary

FROM: Abigail Philip, Cassie Pickens Jewell, Michelle Lee, and Janine Bologna

DATE: 5/18/2018

PAGE: 10

BFWSSI_DRAFT1301	SSI Beneficiaries with BFWSSI_DRAFT in bin	BFWSSI_DRAFT_BOTH1301 compared to BFWSSI_DRAFT1301 (row percent)		
		Less	Same	More
\$50 to \$99	41,736	6.2	93.5	0.5
\$100 to \$199	53,965	6.1	93.3	0.5
\$200 to \$299	34,121	5.1	94.5	0.4
\$300 to \$399	25,873	3.6	96.1	0.3
\$400 to \$499	24,164	3.4	95.9	0.7
\$500 to \$599	11,761	4.6	95.1	0.4
\$600 and above	64,504	16.9	83.1	0.0

Note: BFWSSI_DRAFT is the variable in the DAF16 Annual file, BFWSSI_DRAFT_BOTH is the revised version in the DAF16 SCWF that takes FCI and BMF into account. Note that in 2013, FBR was \$710.

¹ The "same" value in this row includes cases that have missing or zero BFWSSI_DRAFT_BOTH; users of the SCWF can assume that when BFWSSI_DRAFT_BOTH is missing and STWSSI_FCI is non-missing, that the beneficiary had no BFWSSI_DRAFT_BOTH in the month.

- ***“DAF16 data reflects the DCF, with no special handling instituted as a result of this change in Ticket mailings. In the DCF table we use, the TKTMAILDTn is set by default to the TKTSELDTn for beneficiaries who became eligible for the Ticket program between June 2011 and March 2015. The TKTSELDTn and TKTMAILDTn data variables are populated to the same extent as the underlying data source. So while the TKTMAILDTn remains in the DAF, it will show a date that is the same as that of the TKTSELDTn between June 2011 and March 2015, even when no Ticket was mailed. For tickets mailed from April 2015, it will show the actual mail date.”*** I do not think this is correct. We did make changes to how we record ticket mailings in the DAF16 to record the type of ticket sent: initial, 1-year anniversary, and 3-year anniversary. Per Dawn: Looking at the mailing activity since 2015 you should use the variables TKTREMLDDTn and TKTMAILTYPn. The TKTREMLDDTn date will tell you when a Ticket mail activity occurred and the similarly n-suffixed TKTMAILTYPn will tell you what kind of mailing it was (initial (0001), adhoc re-mail (2001), one-year anniversary (3001), three-year anniversary (4001), or catch-up(5001)). Also OBOES told me the interpretation of Mail Date changed in 2015 (per Terry Hetter 3/29/2017): TKT column MAILD_DT no longer reflects the mailed date. Mailed date is set equal to selected date when a ticket is created. You should not see large differences between selected date and mailed date for tickets where the selection date is later than 12/31/2012. Mailing information may be on TKTSENT. This table was created about the same time as the start of the first and third year anniversary notices.

- We have updated the text to capture the changes to how we record ticket mailings in the DAF16:

MEMO TO: Paul O'Leary
FROM: Abigail Philip, Cassie Pickens Jewell, Michelle Lee, and Janine Bologna
DATE: 5/18/2018
PAGE: 11

“The DAF16 incorporates the data from multiple SSA source files to populate the TKTMAILDTn, reflecting the various ticket mailing schemes that have been in place over time. The initial mailings, those through 2012, exist on one DCF table, TKT. Additionally, mailings that occurred during the “hiatus” when SSA did not automatically send tickets are recorded in the TKT table. Mailings occurring after the hiatus, or those in 2015 or after, are carried on another DCF table, TKTSENT. Anniversary mailings are not reflected in the TKTMAILDTn variable, but can be found in the TKTSENT table by using the variable TKTREMLDDTn with TKTMAILTYPn (which indicates the type of mailing). Beneficiaries who do not appear on the TKTSENT table, those that have not had a new initial, catch-up, or anniversary mailing, but do appear on the TKT table will have a TKTMAILDTn that reflects the mail date appearing on the underlying data from the TKT table. If that date occurs between 2013 and March 2015 then the TKTMAILDTn value reflects the date the bene was selected for participation in the TTW program although no physical ticket was mailed.” (p. 65)

Volume 7: DAF16 Development History and Construction Methods

- **Can you please clarify the criteria you use to determine whether a claims history is complete? Can you also provide some additional justification for why you think it is appropriate to discard these incomplete claims? Maybe include this additional information in a footnote.³**
 - We have updated the text in this section as well as included a footnote.

“Many SSNs have multiple VRRMS claims records; for multiple claims records, we determine if the claims history is complete and discard it if not. For complete histories, we roll all claims records on a spell up to a single record per SSN, summing the payment amounts. For records with only a single claim, no summing of payments is needed. We output the SSN-level record keeping all claims, a maximum of 9 in DAF16, in n-suffixed variables, following traditional DAF naming conventions.” (p. 15)

Footnote: “Incomplete claims are identified by examining the earliest available claim detail record within a claim (lowest value of C_DETL_CNTR within a claim). If that claim detail record does not indicate that it is the first such record for the claim (C_DETL_CNTR ne 1) then we delete the entire claim. If a claim is incomplete we will not be able to determine the appropriate payment amounts after adjustments, corrections, and re-computes.” (p. 15)

³ This comment is from the client’s review of DAF15 documentation and was addressed in DAF16.